1 Problemset 0 - Linear Algebra and Multivariable Calculus

1.1 Gradients and Hessians

Recall that a matrix $A \in M_{n\times n}(\mathbb{R})$ is *symmetric* if and only if $A^T = A$. Also, the gradient vector of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as:

$$\nabla f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{bmatrix}$$

Where $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$.

The hessian $\nabla^2 f(x)$ is the $n \times n$ matrix:

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2}{\partial x_1^2} f(x) & \frac{\partial^2}{\partial x_1 x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_1 x_n} f(x) \\ \frac{\partial^2}{\partial x_2 x_1} f(x) & \frac{\partial^2}{\partial x_2^2} f(x) & \cdots & \frac{\partial^2}{\partial x_2 x_n} f(x) \\ & & \cdots & \\ \frac{\partial^2}{\partial x_n x_1} f(x) & \frac{\partial^2}{\partial x_n x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_n^2} f(x) \end{bmatrix}$$

Part (a): Let $f(x) = \frac{1}{2}x^T Ax + b^T x$, where A is a symmetric matrix and $b \in \mathbb{R}^n$ is a vector. Compute $\nabla f(x)$.

Solution:

Let $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$, $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & & \cdots & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$ and $b = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$

By some calculations, we obtain $f(x) = \frac{1}{2}(\sum_{i=1}^n \sum_{j=1}^n a_{ij}x_i x_j) + \sum_{i=1}^n b_i x_i$

$$\implies \frac{\partial f}{\partial x_k} = \frac{1}{2}\left(\sum_{j=1}^n a_{kj}x_j + a_{jk}x_j\right) + b_k$$

(Since we remove exactly one x_k from all terms $a_{ij}x_i x_j$ such that either $i = k$ or $j = k$)
Note that $a_{kj} = a_{jk}$ since A is symmetric, we obtain:

$$\frac{\partial f}{\partial x_k} = \sum_{j=1}^n a_{kj}x_j + b_k$$

Hence:

$$\nabla f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^n a_{1j}x_j + b_1 \\ \vdots \\ \sum_{j=1}^n a_{nj}x_j + b_n \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^n a_{1j}x_j \\ \vdots \\ \sum_{j=1}^n a_{nj}x_j \end{bmatrix} = Ax + b$$

Part (b): Suppose $f(x) = g(h(x))$, where $g : \mathbb{R} \to \mathbb{R}$ is differentiable and $h : \mathbb{R}^n \to \mathbb{R}$ is differentiable. Compute $\nabla f(x)$ in terms of h, g.

Solution:
By the chain rule:

$$\frac{\partial f}{\partial x_i} = \frac{\partial f}{\partial h} \cdot \frac{\partial h}{\partial x_i} = g'(h) \cdot \frac{\partial h}{\partial x_i}$$

Therefore:

$$\nabla f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) \\ \frac{\partial}{\partial x_2} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{bmatrix} = g'(h) \cdot \begin{bmatrix} \frac{\partial}{\partial x_1} h(x) \\ \frac{\partial}{\partial x_2} h(x) \\ \vdots \\ \frac{\partial}{\partial x_n} h(x) \end{bmatrix} = g'(h) \cdot \nabla h(x)$$

Part (c): Let $f(x) = \frac{1}{2}x^T Ax + b^T x$, where A is a symmetric matrix and $b \in \mathbb{R}^n$ is a vector. Compute $\nabla^2 f(x)$.

Solution:
Using **part (a)**, we have:

$$\frac{\partial f}{\partial x_i} = \sum_{j=1}^n a_{ij}x_j + b_i$$

Thus:

$$\frac{\partial^2 f}{\partial x_i x_j} = a_{ij}$$

Hence:

$$\nabla^2 f(x) = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & & \vdots & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = A$$

Part (d): Let $f(x) = g(a^T x)$, where $g : \mathbb{R} \to \mathbb{R}$ is continuously differentiable and $a \in \mathbb{R}^n$ is a vector. What are $\nabla f(X)$ and $\nabla^2 f(x)$?.

Solution: Let $a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$ and $h : \mathbb{R}^n \to \mathbb{R}$ such that $h(x) = a^T x, \forall x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$.

By the chain rule:

$$\frac{\partial f}{\partial x_i} = \frac{\partial f}{\partial h} \cdot \frac{\partial h}{\partial x_i} = g'(a^T x) a_i$$

And:

$$\frac{\partial^2 f}{\partial x_i x_j} = g''(a^T x) a_i a_j$$

Thus:

$$\nabla f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) \\ \frac{\partial}{\partial x_2} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{bmatrix} = g'(a^T x) \cdot a$$

$$\nabla^2 f(x) = g''(a^T x) \begin{bmatrix} a_1^2 & a_1 a_2 & \cdots & a_1 a_n \\ a_2 a_1 & a_2^2 & \cdots & a_2 a_n \\ & & \vdots & \\ a_n a_1 & a_n a_2 & \cdots & a_n^2 \end{bmatrix} = g''(a^T x)(a \cdot a^T)$$

1.2 Positive definite matrix

Part (a): Let $z \in \mathbb{R}^n$ be an $n-$vector. Show that $A = zz^T$ is positive semidefinite.

Solution:

Let $z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}$, then $A = zz^T = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \begin{bmatrix} z_1 & z_2 & \cdots & z_n \end{bmatrix} = \begin{bmatrix} z_1^2 & z_1 z_2 & \cdots & z_1 z_n \\ z_2 z_1 & z_2^2 & \cdots & z_2 z_n \\ & & \vdots & \\ z_n z_1 & z_n z_2 & \cdots & z_n^2 \end{bmatrix}$

It is clear that A is symmetric, hence the quadratic form of A is equal to:

$$\begin{aligned} x^T A x &= \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j \\ &= \sum_{i=1}^{n} A_{ii} x_i^2 + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} A_{ij} x_i x_j \\ &= \sum_{i=1}^{n} (z_i x_i)^2 + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} (z_i x_i)(z_j x_j) \\ &= \Big(\sum_{i=1}^{n} z_i x_i \Big)^2 \geq 0, \forall x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n \end{aligned}$$

The above inequality implies that A is positive semidefinite (Q.E.D).

Part (b): Let $z \in \mathbb{R}^n$ be a *non-zero* $n-$vector and $A = zz^T$. What is the null-space of A? What is the rank of A?

Solution:

We consider two cases:
Case 1: There exists $i \in \{1, ...n\}$ such that $z_i \neq 0$. Without lack of generality, suppose $z_1 \neq 0$:
Applying row operations (note that $z_1 \neq 0$):

$$A = \begin{bmatrix} z_1^2 & z_1 z_2 & \cdots & z_1 z_n \\ z_2 z_1 & z_2^2 & \cdots & z_2 z_n \\ & & \vdots & \\ z_n z_1 & z_n z_2 & \cdots & z_n^2 \end{bmatrix} \Longleftrightarrow \begin{bmatrix} z_1^2 & z_1 z_2 & \cdots & z_1 z_n \\ z_2 z_1^2 & z_1 z_2^2 & \cdots & z_1 z_2 z_n \\ & & \vdots & \\ z_n z_1^2 & z_1 z_n z_2 & \cdots & z_1 z_n^2 \end{bmatrix}$$

$$\Longleftrightarrow \begin{bmatrix} z_1^2 & z_1 z_2 & \cdots & z_1 z_n \\ 0 & 0 & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & 0 \end{bmatrix} \Longleftrightarrow \begin{bmatrix} 1 & \frac{z_2}{z_1} & \cdots & \frac{z_n}{z_1} \\ 0 & 0 & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

Since the last matrix is in reduced row-echelon form, we conclude that nullspace of A is the set $N = \{x \in \mathbb{R}^n, z \cdot x = 0\}$ and rank of the matrix is 1.

Case 2: $z_i = 0, \forall i \in \{1, ...n\}$. In this case, nullspace of A is $\mathbb{R}^n$ and rank of the matrix is 0 (since A is a matrix full of zeros).

Part (c): Let $A \in M_{n\times n}(\mathbb{R})$ be positive semidefinite and $B \in M_{m\times n}(\mathbb{R})$ be arbitrary, where $m, n \in \mathbb{N}$. Is BAB^T PSD? If so, prove it. If not, give a counter-example with explicit A, B.

Solution:

Claim: $C = BAB^T$ is symmetric matrix, provided that A must be symmetric.
Indeed, we have: $(BAB^T)^T = BA^T B^T = BAB^T$.
Consider the quadratic form of C :

$$x^T C x = x^T (BAB^T) x = (x^T B) A (B^T x), x \in \mathbb{R}^m$$

Let $u = B^T x \in \mathbb{R}^n$, then since A is PSD, we have $x^T (BAB^T) x = u^T A u \geq 0, \forall u \in \mathbb{R}^n$.
This also implies that $x^T (BAB^T) x \geq 0, \forall x \in \mathbb{R}^m$ which means BAB^T is PSD.

1.3 Eigenvectors, eigenvalues, and the spectral theorem

Notation: In this section, we use the notation $diag(\lambda_1, \lambda_2, ..., \lambda_n)$ to denote the diagonal matrix with entries $\lambda_1, ..., \lambda_n$, that is:

$$diag(\lambda_1, ..., \lambda_n) = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

Part (a): Suppose that $A \in M_{n\times n}(\mathbb{R})$ is diagonalizable, which means there exists a diagonal matrix $D = diag(\lambda_1, ..., \lambda_n) \in M_{n\times n}(\mathbb{R})$ and $P \in M_{n\times n}(\mathbb{R})$ such that $D = P^{-1}AP$. Let $p^{(i)}, i = \overline{1, ...n}$ be column vectors of P. Show that $Ap^{(i)} = \lambda_i p^{(i)}$, so that the eigenvalues-eigenvectors pairs of A are $(\lambda_i, p^{(i)})$.

Solution:

Since $D = P^{-1}AP$, we have $PD = AP \iff DP = AP$ (since D is diagonal so $PD = DP$) Thus:

$$\begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} p^{(1)} & p^{(2)} & \cdots & p^{(n)} \end{bmatrix} = \begin{bmatrix} Ap^{(1)} & Ap^{(2)} & \cdots & Ap^{(n)} \end{bmatrix}$$

Multiplying in, we obtain:

$$\begin{bmatrix} \lambda_1 p^{(1)} & \lambda_2 p^{(2)} & \cdots & \lambda_n p^{(n)} \end{bmatrix} = \begin{bmatrix} Ap^{(1)} & Ap^{(2)} & \cdots & Ap^{(n)} \end{bmatrix}$$

Equating the corresponding column vectors, we have $Ap^{(i)} = \lambda_i p^{(i)}, \forall i \in \{1, ...n\}$ (Q.E.D).

Recall: A matrix U is called orthogonal if $U^T U = I$. **The Spectral Theorem** states that if $A \in M_{n\times n}(\mathbb{R})$ such that $A = A^T$, then A is *diagonalizable* by an orthogonal matrix P, in other words, $D = P^T AP$ for some diagonal matrix D and orthogonal matrix P.

Part (b): Let $A \in M_{n\times n}(\mathbb{R})$ be symmetric. Show that if $U = \begin{bmatrix} u^{(1)} & u^{(2)} & \cdots & u^{(n)} \end{bmatrix}$ is orthogonal, where $u^{(i)} \in \mathbb{R}^n$ and $A = UDU^T$, then $u^{(i)}$ is an eigenvector of A and λ_i is the corresponding eigenvalue, provided that $D = diag(\lambda_1, ..., \lambda_n)$.

Solution:

Note that $U^T = U^{-1}$, we can apply the proof of **Part (a)** to yield the desired result.

Part (c): Show that if $A \in M_{n\times n}(\mathbb{R})$ is positive semidefinite, then all eigeinvalues of A is greater than or equal to 0.

Solution:

Since $A \in M_{n\times n}(\mathbb{R})$ is positive semidefinite, it must be symmetric and hence diagonalizable by an orthogonal matrix P (**Spectral Theorem**).
In other words, there exists $D = diag(\lambda_1, ..., \lambda_n)$ (λ_i's are eigenvalues of A) such that $D = P^T AP$.
$\forall x \in \mathbb{R}^n$, let $y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \in \mathbb{R}^n$ such that $x = Py$ (or $y = P^{-1}x$), then the quadratic form of A can be rewritten as:

$$Q_A(x) = x^T Ax = y^T P^T APy = y^T Dy = \sum_{i=1}^{n} \lambda_i y_i^2, \forall x \in \mathbb{R}^n$$

And since for each $x \in \mathbb{R}^n$, there exists one corresponding $y = P^{-1}x$, so:

$$Q_A(x) = Q'_A(y) = \sum_{i=1}^{n} \lambda_i y_i^2, \forall y \in \mathbb{R}^n$$

Since A is PSD, this implies that:

$$\sum_{i=1}^{n} \lambda_i y_i^2 \geq 0, \forall y \in \mathbb{R}^n$$

Suppose there exists $j \in \{1, ...n\}$ such that $\lambda_j < 0$, by taking arbitrarily fixed values for x_i's where $i \neq j$ and observe that:

$$\lim_{y_j \to \infty} \sum_{i=1}^{n} \lambda_i y_i^2 = -\infty$$

Which means we can choose a vector $y \in \mathbb{R}^n$ so that $Q_A(x) = Q'_A(y) < 0$, contradiction.
Therefore all eigenvalues λ_i's of A must be greater than or equal to 0 (Q.E.D).